datagen

Simulating the World for Production AI

About Datagen

Founded	Simulation Experts	Data Focus	Verticals	Customers
2018	85+	Humans in Context	Robotics, Security, Automotive, AR/VR	Tech Giants & Fortune-500 Companies

Academic advisors

Prof. Michael J. Black
Founding Director, MPI-IS

Prof. Trevor Darrell
Founder, BAIR

Prof. Gal Chechik
Prof. @ Bar-Ilan

Prof. Lihi Zelnik
Prof. @ Technion

Anthony Goldbloom
CEO, Kaggle

Prof. Fernando De La Torre
Prof. @ CMU, CEO FacioMetrics

datagen

Computer Vision is changing the world

Automotive

Metaverse

Robotics

Industrial

Security

Retail

datagen

Data is the #1 Bottleneck in AI Development

Manual Collection

- ✓ Limited control
- ✓ Data Bias
- ✓ Privacy compliance

Manual Labelling

- ✓ Human errors
- ✓ Inconsistent
- ✓ Labelling Bias

Manual Processing

- ✓ Human errors
- ✓ Slow
- ✓ No Standardization

Testing

Training

Manual Collection

Months of Iterations

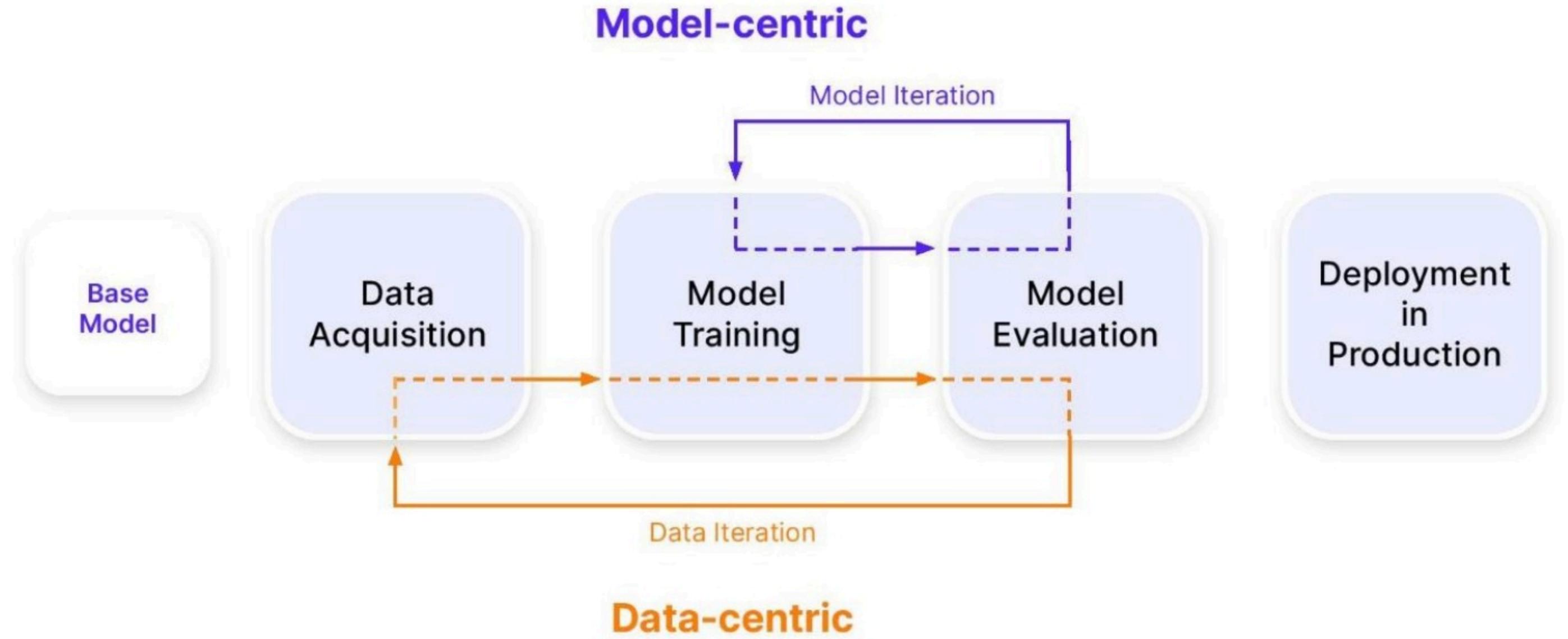Manual Post-processing

Manual Labeling

datagen

Data at Scale is Messy

96% of organizations have problems with training data quality, quantity and speed.

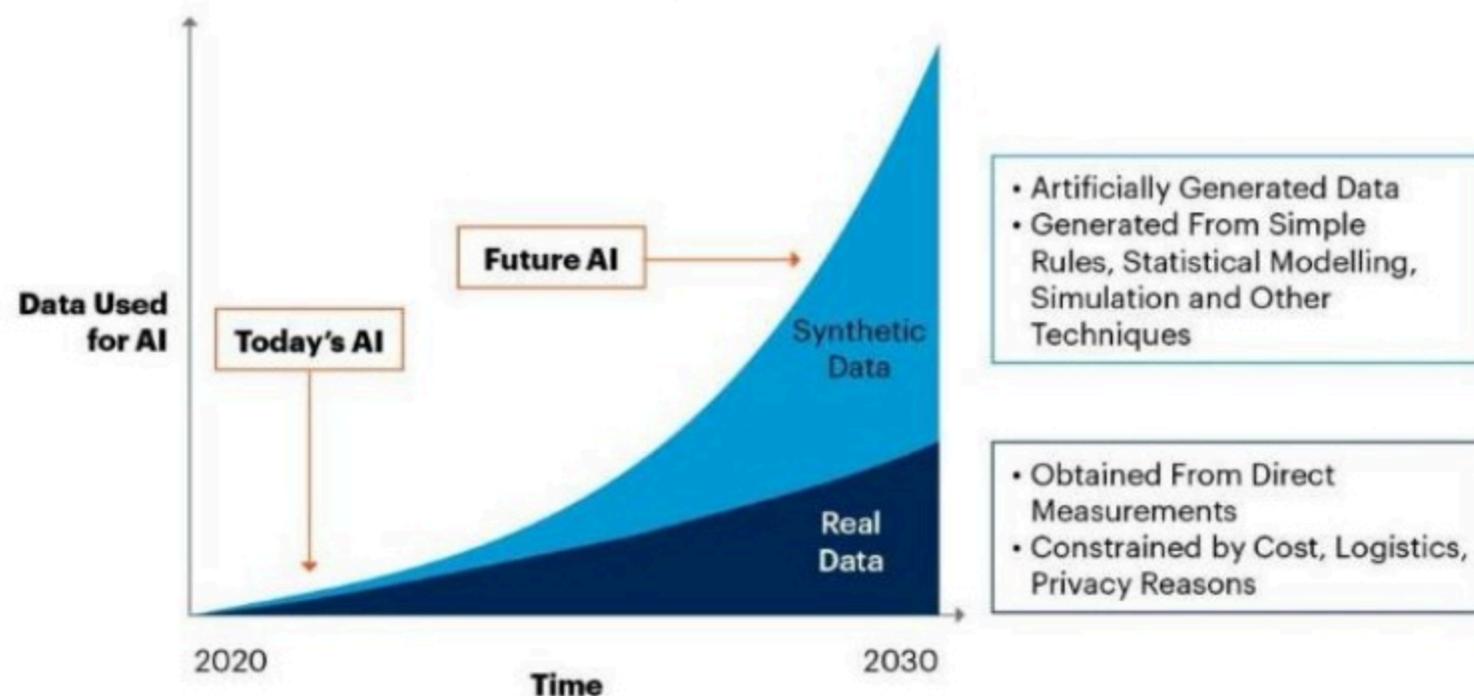Artificial Intelligence and Machine Learning Projects Are Obstructed by Data Issues, May 2019, Dimensional Research

datagen

Data-Centric AI

Synthetic data - paving the way for AI

- Free of Human error
- 2D/3D ground truth
- Granular Control
- Large scale
- Free of Privacy concerns

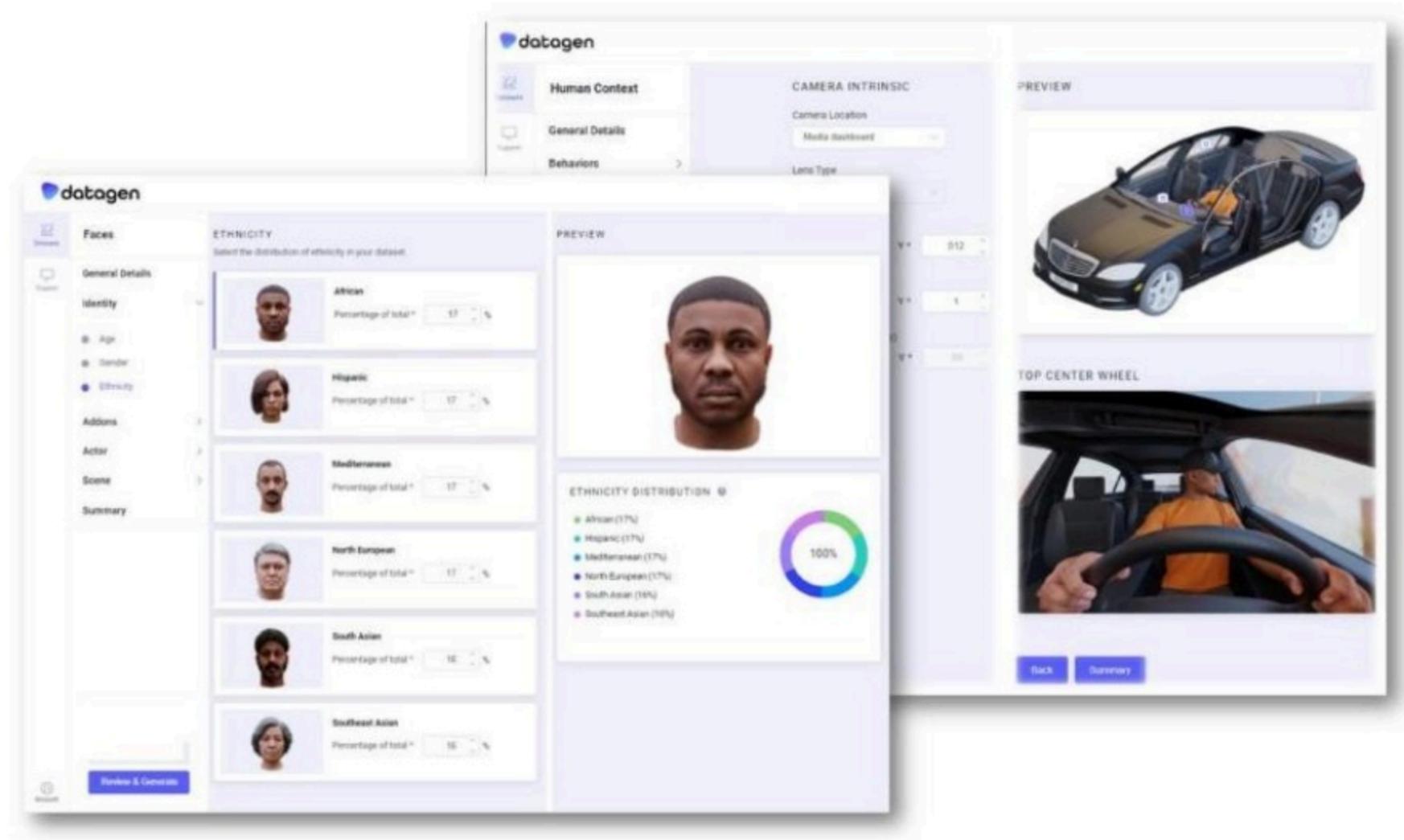**By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models**

Data Used for AI

Today's AI

Future AI

Synthetic Data

Real Data

- Artificially Generated Data
- Generated From Simple Rules, Statistical Modelling, Simulation and Other Techniques

- Obtained From Direct Measurements
- Constrained by Cost, Logistics, Privacy Reasons

2020

Time

2030

Source: Gartner
750175_C

Gartner

datagen

7

The Data Generation Platform

Cloud-based self-service synthetic data generation

- Self-service
- Large scale data generation
- Granular control
- Domain specific

datagen

The Data Generation Platform

Benefits of the platform data

- ✓ High visual domain realism
- ✓ Broad domain diversity
- ✓ Scalable & consistent
- ✓ Pixel perfect & 3D ground truth
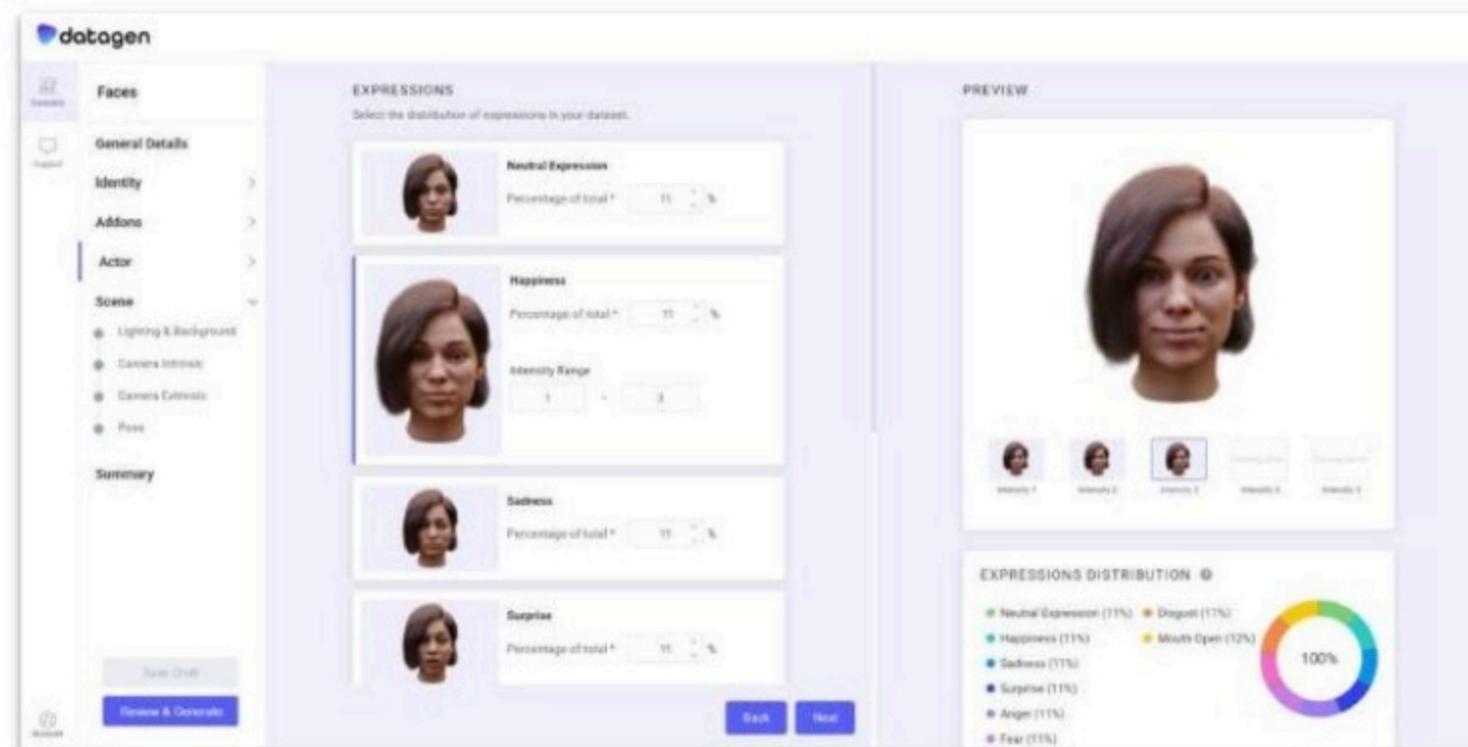- ✓ Fully privacy compliant

datagen

Data Generators

Faces

Humans in Context

Objects in Context
Coming soon

datagen
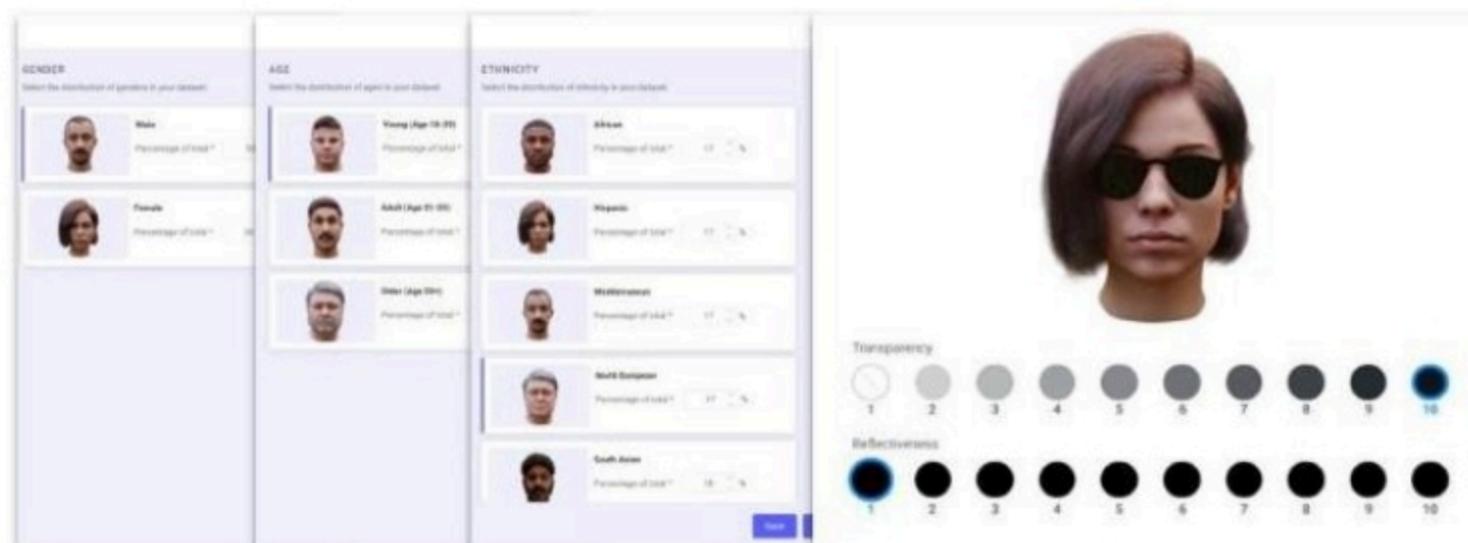
Faces Generator

The platform offers control over

- ✓ Identity - age, gender, ethnicity
- ✓ Scene - camera location, camera intrinsic, lighting & background
- ✓ Eye gaze direction
- ✓ Facial expression
- ✓ Accessories
- ✓ Facial hair
- ✓ Head location and rotation

Computer Vision Tasks

- ✓ Landmark detection
- ✓ Face detection
- ✓ Face recognition
- ✓ Head pose estimation
- ✓ Gaze detection

datagen

11

Faces Generator

Get full control for high variance

Textual Modalities

- ✓ 3D Keypoints (JSON)
- ✓ 3D Dense Keypoints (JSON)
- ✓ 2D Keypoints (JSON)
- ✓ 2D Dense Keypoints (JSON)

Visual Modalities

- ✓ RGB (png)
- ✓ Infrared (png)
- ✓ Depth (exr)
- ✓ Normal (exr)
- ✓ Semantic Segmentation (exr)

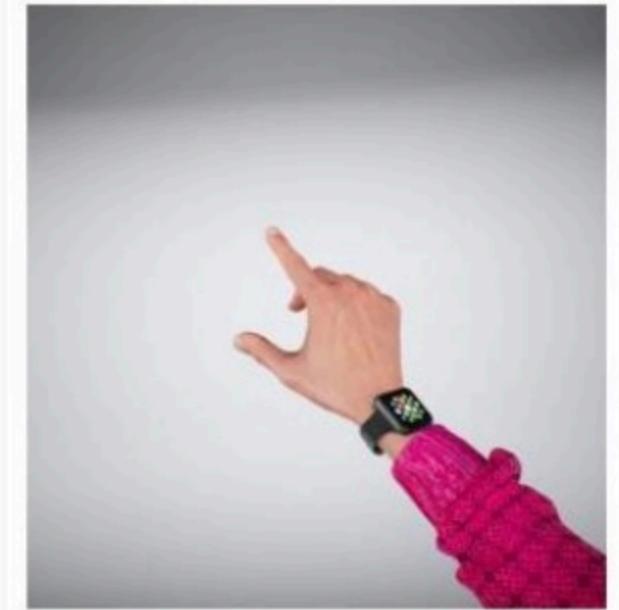**datagen**

Humans in Context (HIC)

Four domains to choose from

In-Cabin Automotive

Smart Office

Home Security

XR/ Metaverse

datagen

HIC Generator for In-Cabin Automotive

Main use cases: Driver Monitoring Systems (DMS) and Occupant Monitoring Systems (OMS)

Computer Vision tasks

- ✓ Visual Distraction
- ✓ Driving Distractions
- ✓ Pose Analysis

datagen

HIC for Home Security

Computer Vision Tasks

✓ Entrance monitoring

✓ Suspicious activity

✓ Package delivery

datagen